REMARKS

Claims 1-83 are pending in this application, of which Claims 1, 16, 37, 45 and 59 are in independent form. Claims 1-83 have been amended as to matters of forms and to define still more clearly what Applicant regards as his invention. A substitute specification, which adds no new matter, is submitted herewith.

Initially, Applicant notes the objection to certain of the claims, requiring that U.S. rather than British spellings be adopted i the claims, at least as regards the word "analysing". As far as Applicant is aware, the only requirement set forth in the Patent Statute and Patent Rules is that applications be filed in English (or that an English translation be submitted after filing). No requirement that U.S. spelling be adopted, is known to Applicant. Nonetheless, to eliminate this as an issue, the claims have been amended to eliminate all non-U.S. spellings, and the same has been done in the specification.

In the Office Action, Claims 33 and 34 were rejected under 35 U.S.C. § 112, second paragraph, as being indefinite. Those claims have been reviewed and amended as deemed necessary to ensure their compliance with all requirements of Section 112, and withdrawal of the rejection under that Section is therefore respectfully requested.

Claims 1, 6/1, 11-16, 24-26, 29-31, 37, 39, 41-45, 51-53, 59 and 63-68 were rejected under 35 U.S.C. § 102(e) as being anticipated by U.S. Patent 6,593,956 (Potts et al.). Claims 2, 3/1, 3/2, 4/1, 4/2, 6/2, 7/1, 7/2, 17, 32, 33, 54, 55, 83/1, 83/16, 83/37, 83/45 and 83/59 were rejected under 35 U.S.C. § 103(a) as being unpatentable over *Potts*, in view of the cited paper "A Statistical Approach To Scene Change Detection" (Sethi et al.). Claims 34-36 and 56-58 were rejected under Section 103(a) as being unpatentable over

Potts in view of *Sethi* and U.S. Patent 6,324,545 (Morag). Claims 5/1, 5/2, 8/1, 8/2, 9/1, 9/2, 10/1, 10/2, 18/1, 18/2, 19/1, 19/2, 20/1, 20/2, 21/1, 21/2, 22/1, 22/2, 23/1, 23/2, 27, 28, 38, 40, 46-50, 60-62 and 69-79 were rejected under Section 103(a) as being obvious from *Potts* in view of *Sethi* and the cited paper "The 'Grammar' of Television and Film" (Chandler). Claims 80-82 were rejected under Section 103(a) as being obvious from *Potts* in view of *Sethi*, *Chandler* and *Morag*.

Independent Claim 1 is directed to a method for automated classification of a digital image, in which the digital image is analyzed for the presence of a human face, and a size of the located face is determined with respect to a size of the image. The digital image is classified according to one of a number of shot types based on the relative size of the face with respect to the image, and the classification of the digital image is stored as metadata associated with the digital image.

In particular, Claim 1 recites that the classification is one according to "one of a number of shot types", and that the classification is stored "as metadata associated with the digital image". These features in combination with the other features of the independent claims, are not taught by either one or any permissible combination of *Potts* and *Sethi*.

In this regard, the purpose of image classification as expressed by *Potts* (see column 10, lines 35 to 62) is to determine whether or not the image contained a human face. While the embodiments described in the present application make use of face detection, the intent of the classification in the method of Claim 1 is not to identify the particular images containing a face, but rather to classify the image based upon a shot type.

As such, while the *Potts* system and the method of Claim 1 both make use of face detection, *Potts* does not teach or suggest classification based on the shot type.

Further, even if *Potts* is deemed to describe storing the location of the face (see column 8, lines 31 to 59), there is no disclosure in *Potts* that the classification of the images are stored as metadata, as recited in Claim 1.

For all these reasons, Claim 1 is believed to be clearly allowable over *Potts*.

Moreover, each of the other independent claims is believed to be allowable over *Potts* taken alone, for the same reasons (at the least) as is Claim 1.

While Applicant agrees that *Sethi* makes numerous references to shot type (on page 4, paragraph 3), *Sethi* fails to provide (despite the contrary statement in paragraph 7) any link between the shot type and the intentions of the photographer. In this regard, page 4, paragraph 3, of *Sethi* needs to be placed into context. In particular, that paragraph is a bullet point relating to one option of "scene change detection" (see the bottom of page 3) which *Sethi* seeks to perform. Specifically, *Sethi* refers to shot type as one test for scene change detection and not, as in the method of Claim 1, for any purpose of classification of an image. Further, in this regard, *Sethi* is silent as to any mechanism of classification of shot type, which is clearly contrary to the intent of the method of Claim 1.

As a consequence, any proper combination of *Potts* and *Sethi* (assuming that any such exists) would result in an arrangement that does not classify images according to shot type as in the method of Claim 1, but rather a system which determines scene changes (*Sethi*) using the detection of faces (*Potts*).

Further, the proposed combination of *Potts* and *Sethi* would still fail to include the express recitation of Claim 1 of the storing of the classification as metadata associated with the digital image.

For all these reasons, it is believed to be clear that Claim 1 (and for the same reasons each of the other independent claims) is allowable over *Potts* and *Sethi*, taken separately or in any permissible combination (if any).

A review of the other art of record has failed to reveal anything which, in Applicants' opinion, would remedy the deficiencies of the art discussed above, as references against the independent claims herein. Those claims are therefore believed patentable over the art of record.

The other claims in this application are each dependent from one or another of the independent claims discussed above and are therefore believed patentable for the same reasons. Since each dependent claim is also deemed to define an additional aspect of the invention, however, the individual reconsideration of the patentability of each on its own merits is respectfully requested.

In view of the foregoing remarks, Applicant respectfully requests favorable reconsideration and early passage to issue of the present application.

Applicant's undersigned attorney may be reached in our New York office by telephone at (212) 218-2100. All correspondence should continue to be directed to our below listed address.

Respectfully submitted,

Leonard P. Diana

Attorney for Applicant

Registration No. 29,296

FITZPATRICK, CELLA, HARPER & SCINTO
30 Rockefeller Plaza
New York, New York 10112-3801
Facsimile: (212) 218-2200

NY_MAIN 404268 v1

TITLE

VISUAL LANGUAGE CLASSIFICATION SYSTEM

Technical Field of the Invention

[0001] The present invention relates generally to the classification of image data and, in particular, to a form of automated classification that permits an editor to automatically generate emotive presentations of the image data.

BACKGROUND

[0002] The editing of video of sequences of images ([[eg.]]e.g., films, video, slide shows), to achieve a desired reaction from an audience traditionally requires input from a human editor who employs techniques other than the mere sequencing of images over a time line. To achieve an understanding by the audience of the intended message or purpose of the production, the editor must draw upon human interpretation methods which are then applied to moving or still images that form the sequence.

[0003] Film makers use many techniques to obtain a desired meaning from images, such techniques including the identification and application of different shot types, both moving and still, the use of different camera angles, different lens types and also film effects. The process of obtaining meaning from the images that make up the final production commences with a story or message that is then translated into a storyboard that is used by the film crew and film director as a

template. Once the film is captured, the editor is then given the resulting images and a shot list for sequencing. It is at an early stage of production, when the screen writer translates the written story or script to a storyboard, that written language becomes visual language. This occurs due to the method by which the audience is told the story and must interpret the message. The visual nature of a moving image generally only has dialogue relevant to the character's experience and, in most cases, is absent of explicit narrative relative to the story being told and the emotional state of the characters within the story. The screen writers must therefore generate this additional information using the visual language obtained from different shot types.

[0004] Examples of different shot types or images are seen in Figs. 1A to 1G. Fig. 1A is representative of an extreme long shot (ELS) which is useful for establishing the characters in their environment, and also orientating the audience as to the particular location. Fig. 1B is representative of a long shot (LS) which is also useful for establishing the characters in their environment and orientating the audience as to the location. In some instances, an ELS is considered more dramatic than the LS. Fig. 1C is representative of a medium long shot (MLS) in which the characters are closer to the viewer and indicates, in a transition from a long shot, subjects of importance to the story. Typically for human subjects, an MLS views those subjects from the knees upwards. Fig. 1D is indicative of a medium shot (MS) in which human characters are generally shown from the waist upwards, and the shot assists the viewer interpreting the characters reactions to their environment and any particular dialogue taking place. Fig. 1E is indicative of a medium closeup (MCU) in which human characters are generally shown from the chest upwards. The MCU is useful for dialogue and communication interpretation including the emotion of the speaking characters. Fig. 1F is indicative of a closeup (CU) which for human characters frames the forehead and shoulders within the shot, and is useful for clear understanding of the emotions associated with any particular dialogue. The closeup is used to consciously place the audience in the position of the character being imaged to achieve a greater dramatic effect. Fig. 1G is representative of an extreme closeup (ECU) formed by a very tight shot of a

portion of the face and demonstrates beyond the dialogue the full dramatic effect of intended emotion. An ECU can be jarring or threatening to the audience in some cases and is often used in many thriller or horror movies. It will further be apparent from the sequence of images in Figs. 1A to 1G that different shots clearly can display different meaning. For example, neither of Figs. 1F and 1G indicate that the subject is seen flying a kite, nor do Figs. 1D or 1E place the kite flying subject on a farm indicated by the cow seen in Figs. 1A to 1C. Further, it is not apparent from Fig. 1A that the subject is smiling or indeed that the subject's eyes are open.

[0005] A photograph or moving image of a person incorporating a full body shot will be interpreted by the viewer as having a different meaning to a shot of exactly the same person, where the image consists of only a closeup of the face of the subject. A full-length body shot is typically interpreted by a viewer as informative and is useful to determine the sociological factors of the subject and the relationship of the subject to the particular environment.

[0006] An example of this is illustrated in Figs. 2A to 2C which show the same subject matter presented with three different shot types. Fig. 2A is a wide shot of the subject within the landscape and is informative as to the location, subject and activity taken close within the scene. Fig. 2B is a mid-shot of the subject with some of the surrounding landscape, and changes the emphasis from the location and activity to the character of the subject. Fig. 2C provides a closeup of the subject and draws the audience to focus upon the subject.

[0007] Panning is a technique used by screen writers to help the audience participate in the absorption of information within a scene. The technique is commonly used with open landscapes or when establishing shots are used in movie productions. A straight shot, obtained when the camera does not move, contrasts the effectiveness of a pan. With a straight shot, the viewer is forced to move their eyes around the scene, searching for information, as opposed to how the pan feeds information to the viewer thus not requiring the viewer to seek out a particular message. The movement of the camera within a pan directs the audience as to those elements within a scene that should be observed and, when used correctly, is

intended to mimic the human method of information interpretation and absorption. Fig. 3A is an example of a still shot including a number of image elements [[(eg.]] (e.g., the sun, the house, the cow, the person and the kite) which the audience may scan for information. In film, a still shot is typically used as an establishing shot so as to orientate the audience with the location and the relationship to the story. The screen writer relies upon this type of shot to make sense of any following scenes. Fig. 3B demonstrates an example of a panning technique combined with a zoom, spread amongst four consecutive frames.

[0008] Further, differing camera angles, as opposed to direct, straight shots, are often used to generate meaning from the subject, such meaning not otherwise being available due to dialogue alone. For example, newspaper and television journalists often use altered camera angles to solicit propaganda about preferred election candidates. For example, interviews recorded from a low angle present the subject as superior to the audience, whereas the presentation of the same subject may be altered if taken from a high angle to give an inferior interpretation. The same technique is commonly used in movie making to dramatically increase the effect of an antagonist and [[their]] his victim. When the victim is shot from a high angle, [[they]] he or she not only appears as weak and vulnerable, but the audience ~~empathises~~ emphathizes with the character and also experiences [[their]] the character's fear.

[0009] Fig. 4A is indicative of an eye level shot which is a standard shot contrasting with angles used in other shots and seen in Figs. 4B to 4E. Fig. 4B shows a high angle shot and is used to place the subject in an inferior position. Fig. 4C is indicative of a low angle shot where the camera angle is held low with the subject projecting them as superior. Fig. 4D is indicative of an oblique angle shot where the camera is held ~~off-centre~~ off-center influencing the audience to interpret the subject as out of the ordinary, or as unbalanced in character. Fig. 4E is representative of a Dutch angle shot which is often used to generate a hurried, "no time to waste" or bizarre effect of the subject. The audience is conveyed a message that something has gone astray in either a positive or negative fashion.

[0010] There are many other types of images or shots in addition to those discussed above that can give insight to the particular story being presented. Tracking shots follow the subject allowing the audience the experience of being part of the action. Panning gives meaning and designates importance to subjects within a scene as well as providing a panoramic view of the scene. A "swish" pan is similar, however is used more as a transition within a scene, quickly sweeping from one subject to another, thus generating a blurred effect. Tilt shots consist of moving the camera from one point up or down, thus mimicking the way in which humans evaluate a person or vertical object absorbing the information presented thereby. A hand-held shot portrays to the audience that the filming is taking place immediately, and if often used to best effect when associated with shots taken when the camera is supported ([[eg.]]e.g., using a tripod or boom).

[0011] To understand the impact visual language has on presenting images in a more meaningful way, it is appropriate to compare the results of contemporary motion pictures with earlier attempts of film making. Early examples of motion pictures consisted of full shots of the characters from the feet upwards reflecting the transition from stage acting. For example, the Charlie Chaplin era of film making and story telling contrasts sharply with later dramatic, emotion filled motion pictures. Pioneering director D.W. Griffiths notably first introduced the use of a pallet of shot types for the purpose of creating drama in film. This arose from a desire of the audience to explore the emotional experience of the characters of the film.

[0012] Film makers also use other techniques to tell their story, such techniques including the choice of lens and film effects. These are all used to encourage the audience to understand the intended message or purpose of the production. The audience does not need to understand how, or even be aware that, these techniques have been applied to the images. In fact, if applied properly with skill, the methods will not even be apparent to the audience.

[0013] The skill required by the successful film maker is typically only acquired through many years of tuition and practice as well as through the collaboration of many experts to achieve a successfully crafted message. Amateur film makers and

home video makers in contrast often lack the skill and the opportunity to understand or employ such methods. However, amateur and home film makers, being well exposed to professional film productions have a desire for their own productions to be refined to some extent approaching that of professional productions, if not those of big-budget Hollywood extravaganzas. Whilst there currently ~~exists~~ exist many film schools that ~~specialise~~ specialize in courses to educate potential film makers with such techniques, attendance at such courses is often prohibitive to the amateur film maker. Other techniques currently available that may assist the amateur film maker typically ~~includes~~ include software products to aid in the sequencing of images and/or interactive education techniques for tutoring prospective film makers. However, current software approaches have not been widely adopted due to prohibitive costs and skill required for use being excessive for small (domestic) productions.

[0014] Time is also a major factor in respect to the current techniques of film editing to unskilled editor. Typically, the time taken to plan shots and their sequencing is substantial and is typically out of the realistic scope of an average home/amateur film maker.

[0015] It is therefore desirable to provide a means by which unskilled (amateur) movie makers can create visual productions that convey a desired emotive effect to an audience without a need for extensive planning or examination of shot types.

SUMMARY OF THE INVENTION

[0016] This need is addressed through the automated classification of images and/or shots into various emotive categories thereby permitting editing to achieve a desired emotive effect.

[0017] According to a first aspect of the present disclosure, there is provided a method for automated classification of a digital image, ~~said method~~ comprising the steps of:

~~analysing said~~ analyzing the image for the presence of a human face;

determining a size of the located face with respect to a size of ~~said~~ the image; and

classifying ~~said~~ the image based on the relative size of ~~said~~ the face with respect to ~~said~~ the image.

[0018] According to a second aspect of the present disclosure, there is provided a method for automated classification of a digital image, ~~said method~~ comprising the steps of:

~~analysing said~~ analyzing the image for the presence of a human face;

determining a position of the located face with respect to a frame of ~~said~~ the image; and

classifying ~~said~~ the image based on the relative position of ~~said~~ the face with respect to ~~said~~ the image frame.

[0019] According to another aspect of the present disclosure, there is provided apparatus for implementing any one of the aforementioned methods.

[0020] According to another aspect of the present disclosure there is provided a computer program product including a computer readable medium having recorded thereon a computer program for implementing any one of the methods described above.

BRIEF DESCRIPTION OF THE DRAWINGS

[0021] One or more embodiments of the present invention will now be described with reference to the drawings, in which:

[0022] Figs. 1A to 1G depict a number of shot ranges used by film makers;

[0023] Figs. 2A to 2C depict three different shot types used by film makers;

[0024] Figs. 3A and 3B depict the effect of a pan in influencing the emotional state of the viewer;

[0025] Figs. 4A to 4E depict various angled camera shots also used by film makers;

[0026] Fig. 5 is a schematic block diagram representation of an image recording and production system;

[0027] Fig. 6 is a schematic block diagram of a general purpose computer system upon which the disclosed arrangements can be practiced; and

[0028] Fig. 7 is a flow chart depicting the use of templates for video editing.

DETAILED DESCRIPTION INCLUDING BEST MODE

[0029] Fig. 5 shows a schematic representation of an image recording and production system 500 where a scene 502 is captured using an image recording device 504, such as a digital video camera or digital still camera. When the scene 502 is captured by a still camera, typically a sequence of still images is recorded, in effect complementing the sequence of images that might be recorded by a video camera. Associated with the capture of the images is the generation of capture data 506 which is output from the camera 504 and typically comprises image data 506a, video data 506b, audio data 506c and "camera" metadata 506d. The camera metadata 506 represents metadata usually generated automatically by the camera or manually entered by the user of the camera. Such can include image or frame number, a real-time of capture possibly include a date, details regarding camera settings (aperture, exposure, etc.) and ambient information such as light measurements, to name but a few

[0030] Where appropriate, the capture data 504 recorded by the camera 504 is transferred 508 to a mass storage arrangement 510, typically associated with a computing system, whereupon the images are made available via an interconnection 520 to a visual language classification system 522. The classification system 508 generates metadata which is configured for convenient editing by the film maker. The visual language classification system 522 outputs classification data 524, configured as further metadata, which is associated with each image and which may be stored within a mass storage unit 526. The classification data 524 in the store 526 may be output to an editing module 514 which, through accessing the image data via a connection 512 to the store 510, provides for the formation of an edited sequence 528 which may be output to a presentation unit 516 for display via a display unit 518, such as a television display, or storage in a mass storage device 519. In some implementations, the stores 510, 526 and 519 may be integrally formed.

[0031] The classification system 522 performs content analysis to ~~analyse~~ analyze the images residing in the store 510. The analysis performed within the

classification system 522 is configured to provide information about the intention of the photographer at the time of capturing the image or image sequence. Such analysis may comprise the detection of human faces and preferably other visually distinct features including landscape features such as the sky, green grass, sandy or brown earth, or other particular shapes such as motor vehicles, buildings and the like, from the image data. Audio analysis where appropriate can be used to identify specific events within the sequence of images such a person talking, the passing of a motor car, or the crack of a ball hitting a bat in a sports game, such as baseball or cricket, for example. The classification system 522 provides metadata related to or indicative of the content identified within an image sequence, or at the particular image within the sequence.

[0032] One specific example of content analysis that may be applied by classification system 522 is that of face detection, that permits identification and tracking of particular human subjects in images or sequences thereof. An example of a face detection arrangement that may be used in the arrangement of Fig. 5 is that described in US Patent No. 5,642,431-A (Poggio [[et.]]et al.). Another example is that disclosed in Australian Patent Publication No. AU-A-33982/99. Such face detection arrangements typically identify within an image frame a group or area of pixels which are skin ~~coloured~~ colored and thus may represent a face, thereby enabling that group or area, and thus the face, to be tagged by metadata and monitored. Such monitoring may include establishing a bounding box about the height and width of the detected face and thereafter tracking changes or movement in the box across a number of image frames.

[0033] In the sequence of images of Figs. 1A to 1G, the fine content of Figs. 1A and 1B are generally too small to permit accurate face detection. As such, those frames may be classified as non-face images. However in each of Figs. 1C to 1G, the face of the person flying the kite is quite discernible and a significant feature of each respective image. Thus, those images may be automatically classified as face images, such classification being identified as metadata 524 generated by content analysis performed by the classification system 522 and linked or otherwise associated with the metadata 506d provided with the images.

[0034] Further, and in a preferred implementation, the size of the detected face, as a proportion of the overall image size, is used to establish and record the type of shot. For example, simple rules may be established to identify the type of shot. A first rule can be that, where a face is detected, but the face is substantially smaller than the image in which the face is detected, that image may be classified as a far shot. A similar rule is where a face is detected which is sized substantially the same as the image. This may be classified as a close-up. An extreme close-up may be where the face occupies the entire image or where it is substantially the same size as the image but extends beyond the edges of the image.

[0035] In another example, in Fig. 1C, which is a MLS, the face represents about 2% of the image. In Fig. 1D, the face occupies about 4% of the image, this being a MS. For Fig. 1E, a MCU delivers the face at a size of about 10% of the image. The CU shot of Fig. 1F provides the face at about 60% of the image, and for [[a]] an ECU, the face is in excess of about 80% of the image. A suitable set of rules may thus be established to define the type of shot relative to the subject, whether or not the subject is a face or some other identifiable image structure ([[eg.]]e.g., cow, house, motor vehicle, etc). Example rules are set out below:

Medium Long Shot (MLS)	subject < 2.5% of the image;
Medium Shot (MS)	2.5% < subject < 10% of the image;
Medium Close Up (MCU)	10% < subject < 30% of the image;
Close Up (CU)	30% < subject < 80% of the image; and
Extreme Close Up (ECU)	subject > 80% of the image.

[0036] Where desired, the film maker may vary the rules depending on the particular type of source footage available, or depending on a particular editing effect desired to be achieved.

[0037] Another example of content analysis for classification is camera tilt angle. This can be assessed by examining the relative position of a detected face in the image frame. For example, as seen in Fig. 4A, where the face is detected centrally within the image frame, this may be classified as a eye-level shot. In Fig. 4B, where the subject is positioned towards the bottom of the frame, such may be classified as a high angle shot. the positioning of the detected face may be

correlated with a tiling of the image frame so as to provide the desired classification. Tiles within the frame may be pre-classified as eye-level, high shot, low shot, left side, and right side. The location of the detected face in certain tiles may then be used to determine an average tile location and thus classify the image according to the position of the average face tile. Such an approach may be readily applied to the images of Figs. 4A to 4D.

[0038] The Dutch shot of Fig. 4E may be determined by detecting edges within the image. Such edges may be detected using any one of a large number of known edge detection arrangements. Edges in images often indicate the horizon, or some other horizontal edge, or vertical edges such as those formed by building walls. An edge that is detected as being substantially non-vertical and non-horizontal may thus indicate a Dutch shot. Classification may be performed by comparing an angle of inclination of the detected edge with the image frame. Where the angle is about 0 degrees or about 90 degrees, such may be indicative of ~~an horizon~~ a horizontal or vertical wall, respectively. Such may be a traditional shot. However, where the angle of inclination is substantially between these values, a Dutch shot may be indicated. Preferred angles of inclination for such detection may be between 30 and 60 degrees, but may be determined by the user where desired.

[0039] In an alternative implementation, the visual language classification system can permit the user to supplement the classification with other terms relating to the emotive message conveyed by the scene. Such manually entered metadata may include terms such as "happy", "smiling", "leisure", and "fun" in the example of Figs. 1C to 1G. More complicated descriptions may also be entered, such as "kite flying". This manually enter metadata that can supplement the automatically generated metadata and be stored with the automatically generated metadata.

[0040] As a result of such processing, the store 526 is formed to include metadata representative of the content of source images to be used to form the final production. The metadata not only includes timing and sequencing ([[eg.]]e.g., scene number, etc.) information, but also information indicative of the content of the images and shot types which can be used as prompts in the editing process to follow.

[0041] With the database 526 formed, the user may then commence editing the selected images. This is done by invoking an editing system 514 which extracts the appropriate images or sequence of images from the store 510. Using the information contained within the metadata store 526, the user may conveniently edit particular images. The database information may be used to define fade-in and fade-out points, images where a change in zoom is desired, points of interest within individual images which can represent focal ~~centres~~ centers for zooming operations either or both as source or target, amongst many others.

[0042] Editing performed by the editing system 514 may operate using the classifications 524 in a variety of ways. For example, the user may wish to commence an image sequence with a long shot, and hence may enter into the system 514 a request for all long shots to be listed. The system 514 then interrogates the store 526 to form a pickiest of images that have been previously classified as a long shot. The user may then select a long shot from the list to commence the edited sequence. The classification thus substantially reduces the user's editing time by providing a ready source of searchable information regarding each image or shot sequence. Another example is where the user wishes to show the emotion "fear" in the faces of the subjects. Since faces are typically not detected in any significant detail for anything under a medium shot, a search of the store 526 may be made for all medium shots, close-ups and extreme close-ups. A corresponding pick list results from which the user can conveniently review a generally smaller number of images than the total number available to determine those that show "fear". User entered metadata such as "fear" may then supplement the automatically generated classification for those images that display such an emotion.

[0043] The automated content analysis of images as discussed above permits the rapid processing of sequences of images to facilitate the formation of an enhanced edited result. For example, where a video source is provided having 25 frames per second, a 5 second shot requires the editing of 125 frames. To perform manual face detection and focal point establishment on each frame is time consuming and prone to inconsistent results due to human inconsistency. Through automation by

content analysis, the positions of the face since each frame may be located according to consistently applied rules. All that is then necessary is [[form]] for the user to select the start and end points and the corresponding edit functions ([[eg.]]e.g., zoom values from 0% at the start, and 60% at the end).

[0044] Metadata analysis of the source material may include the following:

(i) time code and date data;

(ii) G.P.S. data;

(iii) image quality analysis (sharpness, ~~colour~~ color, content quality, etc.);

(iv) original shot type detection;

(v) object detection and custom object detection (determined by the author);

(vi) movement detection;

(vii) face detection;

(viii) audio detection;

(ix) collision detection;

(x) tile (interframe structure) analysis; and

(xi) user entered metadata.

[0045] The method described above with reference to Fig. 5 is preferably practiced using a conventional general-purpose computer system 600, such as that shown in Fig. 6 wherein the processes of Fig. 5 may be implemented as software, such as an application program executing within the computer system 600. The software may be divided into two separate parts; one part for carrying out the classification and editing methods, and another part to manage the user interface between the latter and the user. The software may be stored in a computer readable medium, including the storage devices described below, for example. The software is loaded into the computer from the computer readable medium, and then executed by the computer. A computer readable medium having such software or computer program recorded on it is a computer program product. The use of the computer program product in the computer preferably effects an advantageous apparatus for classification and consequential editing of images or sequences of images.

[0046] The computer system 600 comprises a computer module 601, input devices such as a keyboard 60 and mouse 603, output devices including a printer 615 and a visual display device 614 and loud speaker 617. A Modulator-Demodulator (Modem) transceiver device 616 is used by the computer module 601 for communicating to and from a communications network 620, for example connectable via a telephone line 621 or other functional medium. The modem 616 can be used to obtain access to the Internet, and other network systems, such as a Local Area Network (LAN) or a Wide Area Network (WAN).

[0047] The computer module 601 typically includes at least one processor unit 605, a memory unit 606, for example formed from semiconductor random access memory (RAM) and read only memory (ROM), input/output (UO) interfaces including a audio/video interface 607, and an I/O interface 613 for the keyboard 602 and mouse 603 and optionally a joystick (not illustrated), and an interface 608 for the modem 616. A storage device 609 is provided and typically includes a hard disk drive 610 and a floppy disk drive 611. A magnetic tape drive (not illustrated) may also be used. A CD-ROM drive 612 is typically provided as a non-volatile source of data. The components 605 to 613 of the computer module 601, typically communicate via an interconnected bus 604 and in a manner which results in a conventional mode of operation of the computer system 600 known to those in the relevant art. Examples of computers on which the described arrangements can be ~~practised~~ practiced, include IBM-PC's and compatibles, Sun Sparcstations or alike computer systems evolved therefrom.

[0048] Typically, the application program is resident on the hard disk drive 610 and read and controlled in its execution by the processor 605. Intermediate storage of the program and any data fetched from the network 620 may be accomplished using the semiconductor memory 606, possibly in concert with the hard disk drive 610. In some instances, the application program may be supplied to the user encoded on a CD-ROM or floppy disk and read via the corresponding drive 612 or 611, or alternatively may be read by the user from the network 620 via the modem device 616. Still further, the software can also be loaded into the computer system 600 from other computer readable medium including magnetic tape, a ROM or

integrated circuit, a magneto-optical disk, a radio or infra-red transmission channel between the computer module 601 and another device, a computer readable card such as a PCMCIA card, and the Internet and Intranets including e-mail transmissions and information recorded on Websites and the like. The foregoing is merely exemplary of relevant computer readable media. Other computer readable media may also be used.

[0049] The method described with reference to Fig. 6 may alternatively or additionally be implemented in dedicated hardware such as one or more integrated circuits performing the functions or sub functions of the system. Such dedicated hardware may include graphic processors, digital signal processors, or one or more microprocessors and associated memories. For example, specific visual effects such as zoom and image interpolation may be performed in specific hardware devices configured for such functions. Other processing modules, for example, used for face detection or audio processing, may be performed in dedicated DSP apparatus.

[0050] The description above with respect to Fig. 5 indicates how the editing system 514 may be used to create an output presentation based upon classifications derived from the image content. A further approach to editing may be achieved using a template-based approach 700 depicted in the flow chart of Fig. 7, which for example may be implemented within the editing system 514. The method 700 commences at step 702 where a desired clip, being a portion of footage between a single start-stop transition, is selected for processing. A number of clips may be processed in sequence to create a production. This is followed by step 704 where a desired template is selected for application to the clip. A template in this regard is a set of editing rules that may be applied to various shot and clip types to achieve a desired visual effect. Alternatively, a template need only be applied to a portion of a clip, or in some instances one or still images or video extracts for which processing is desired. Typically a number of templates 706 are available for selection 708. Each template 706 may be established as a Boolean set of rules each with a number of default settings. An example template is depicted in Table 1

below and which defines particular visual effects that are top be applied to particular shot types.

Table 1

Template #2		Effect									
Shot type	Select	Speed of replay					B&W	Zoom time	Color filter	Sound	etc.
		x¼	x½	x1	x2	x4					
ECU	1	1					1	0	1	0	
CU	1	1					1	0	1	0	
MSU	1			1			1	+2	1	0	
MS	0										
MLS	0										
LS	0										
Other#1	1					1	1	0	1	1	
Other#2	0										

[0051] In the template of Table 1, the various shot types are listed based upon face detection criteria described above. Two "other" shot types are shown, these for example being where no face is detected or some other detectable event may be determined. Such for example may be frames containing a white ~~coloured~~ colored motor racing car of particular interest to the user, as compared to other ~~coloured~~ colored racing cars that may have been captured. Such a racing car may be detected by the classification system 522 being arranged to detect both a substantial region of the ~~colour~~ color white and also substantial movement of that ~~colour~~ color thereby permitting such frames to be classified as "Other#1". The movement may be actual movement of the racing car across the frame over a series of adjacent frames, or relative movement where the racing car appears substantially stationary within the series of adjacent frames, whilst substantial movement of the background occurs. Such a classification may be formed independent of the ECU, CU, MCU etc. approach described above. As seen from Table 1, each of ECU,

CU, MCU and Other#1 shot types are selected for inclusion in the edited presentation.

[0052] The template ([[ie.]]i.e., template #2) selected 710 may altered according to a user determination made in step 712. Where alteration is desired, step 714 follows which permits the user to modify the Boolean values within the template table. As seen above, those shot types not selected ([[ie.]]i.e., MS, MLS, LS and Other#2) are disabled from the table, as indicated by the shading thereof. Those selected shot types may then have their corresponding effects modified by the user. As shown a number of different speeds of replay are provide, the selection of one for any shot type disabling the others for the same shot type. As seen each of the ECU and CU are selected to replay at quarter speed, whereas the MCU replays at natural speed. The racing car captured by the Other#1 shot type is selected for replay at four times speed to fulfil the user's desire to accentuate the differences between facial and motor car shots. Each of the selected shots has a monochrome (B&W) setting selected, thereby removing ~~colour~~ color variation, although a ~~colour~~ color filter effect has been enabled. Such an effect may provide a constant orange/brown tinge to the entire frame and in this example would result in the images been reproduced with an aged-sepia effect. Sound is seen disabled on the facial shots but enabled on the racing car shots.

[0053] A zoom feature is also provided to permit translations between adjacent shot types. As seen in the example of Table 1, MCU shots are subject to a zoom of "+2", this notation representing a zoom-in to the next shot type ([[ie.]]i.e., CU) with the zoom occurring over a period of 2 seconds. Typically, during the zoom, the image is automatically cropped to retain a size within that of the display. Zoom-outs are also possible and are indicated by a minus symbol (-). Durations may be specified in seconds, frames, or as being instantaneous ([[eg.]]e.g., ++), the later directly creating a new frame for inclusion in the edited production. The transitions for zoom in Table 1 are specified as occurring between adjacent shot types. Alternatively the degree of zoom and the zoom duration may be separately specified for each shot type ([[eg:]]e.g., MCU : 150%: 25 frames; CU : 200%: 10 frames; ECU : 30% : 50 frames). In this fashion, the edited production may show

for a particular shot type a zoom to another shot type over a predetermined period thereby enhancing the emotional effect of the production. For example, a zoom from an MCU to an ECU may form part of a "dramatic" template, being one where ECU's are used to focus the viewer's attention on the central character. A "tribute" template may include a zoom from a MCU to a CU.

[0054] Other types of image editing effects may be applied within a template as desired.

[0055] Once modified, the template is stored and control returns to step 704 where the user may select the template just modified. Once a template has been selected, step 716 follows where the sequence of clips is derived form the camera metadata retained in the store 718. Once the correct sequence is formed, the sequence is edited in step 720 by applying the selected template to the sequence. This step involves sourcing firstly the classification metadata from the store 718 to determine the shots types and then sourcing the video data to which the various effected selected for that shot may be applied. This results [[ion]] <u>in</u> the output presentation of step 722 which may be sent for storage or directly reproduced to a display arrangement.

[0056] It will be appreciated that a variety of templates may be created, each having the capacity to impose on the source image data a particular emotive editing style in response to the classification of shot types contained therein. Further, individual clips or scenes may be edited using different templates thereby altering the presentation style based upon the subject matter. Accordingly, a family visit to the motor races may include scenes depicting a picnic lunch using substantially natural footage but limited to MS's and MLS's, action scenes edited in the manner described above with respect to Table 1, and super-action scenes where substantial slow motion is used to accentuate a crash during the race. The crash may be classified by the user supplementing the metadata of that portion of footage with a tag indicating importance. Also, whilst the template of Table 1 relies predominantly on shot distance, other classifications such as tilt angle as discussed above may alternatively or additionally be included.

INDUSTRIAL APPLICABILITY

[0057] The arrangements described are applicable to the image editing and reproduction industries and find particular application with amateur movie makers who are trained in the intricacies of shot and subject identification, and consequential editing based thereupon.

[0058] The foregoing describes only some embodiments of the present invention, and modifications and/or changes can be made thereto without departing from the scope and spirit of the present invention, the described embodiments being illustrative and not restrictive.

ABSTRACT

VISUAL LANGUAGE CLASSIFICATION SYSTEM

[0059] Disclosed [[is]] are a method and system ~~(500)~~ for automated classification of a digital image ~~(502)~~. The method ~~analyses~~ analyzes the image for the presence of a human face. A determination is then made regarding the size of the located face compared to the size of the ~~image(Figs.1A-1G~~ image to classify the image based on the relative size of the face. Alternatively, the position of the face within the image can be used to determine the classification. With a classified image, particularly forming part of a sequence of classified images, editing ~~(514)~~ of the sequence may be performed dependent upon the classification to achieve a desired aesthetic effect. The editing may be performed with the aid of an editing template ~~(706)~~.